ON A THEOREM OF BEHREND

P. ERDÖS, A. SÁRKÖZY and E. SZEMERÉDI

(Received 15 September 1965)

Dedicated to the memory of Felix Behrend

A sequence of integers $0 < a_1 < a_2 < \cdots$ no term of which divides any other will be called a primitive sequence. Throughout this paper $c_1, c_2, \cdots$ will denote suitable positive absolute constants. Behrend [1] proved that for every primitive sequence

$$f_A(x) = \sum_{a_i < x} \frac{1}{a_i} < c_1 \log x/(\log\log x)^{\frac{1}{2}}. \tag{1}$$

Sivasankaranarayana Pillai observed that (1) is in a sense best possible. He showed that there is a c_2 so that for every x there is a primitive sequence $a_1 < \cdots < a_k \leqq x$ for which

$$f_A(x) > c_2 \log x/(\log\log x)^{\frac{1}{2}}. \tag{2}$$

In the present paper we are going to prove the following

THEOREM 1. *Let A be an infinite primitive sequence. Then*

$$f_A(x) = o\left(\log x/(\log\log x)^{\frac{1}{2}}\right). \tag{3}$$

Our Theorem shows that though for a finite primitive sequence (1) is best possible, it can nevertheless be improved for infinite primitive sequences.

Before proving our Theorem we show that it is best possible. In fact we shall show that if $h(x) \to \infty$ arbitrarily slowly, then there exists a primitive sequence A so that

$$\limsup_{x\to\infty} f_A(x)h(x)\ (\log\log x^{\frac{1}{2}})/\log x = \infty. \tag{4}$$

We only outline the proof of (4) since the details can easily be filled in by the reader using the methods of [3]. Let $x_1 < x_2 < \cdots$ tend to infinity sufficiently fast. In the interval $(x_{\nu-1}, x_\nu)$ our sequence consists of the integers having exactly $[\log\log x_\nu]$ distinct prime factors greater than $x_{\nu-1}$ (and no prime factor $\leqq x_{\nu-1}$). A simple computation shows that if $x_\nu \to \infty$ sufficiently fast (depending on $h(x)$) then (4) holds.

To prove Theorem 1 we assume that there is a primitive sequence $A = \{a_1 < a_2 < \cdots\}$ for which (3) does not hold and we will obtain a contradiction. First of all we observe that if a sequence A exists for which (3) does not hold, we can assume that there is such a sequence all whose terms are squarefree. Put

$$A = \bigcup_{k=1}^{\infty} A^{(k)}$$

where the greatest square factor of the integers of $A^{(k)}$ is k^2. It easily follows from $\sum_{k=1}^{\infty} 1/k^2 < \infty$ and (1) that if A does not satisfy (3) then for some fixed k_0, $A^{(k_0)}$ also does not satisfy (3). Put

$$A^{(k_0)} = \{a_1^{(k_0)} < a_2^{(k_0)} < \cdots\}.$$

Clearly $a_i^{(k_0)} = k_0^2 b_i$ where b_i is squarefree and $b_1 < b_2 < \cdots$ evidently does not satisfy (3).

Henceforth we assume that A is a primitive sequence of squarefree numbers for which (3) does not hold. Then there clearly exists a sequence $x_1 < x_2 < \cdots$ tending to infinity sufficiently fast (this will be specified later) so that

$$\sum_{x_{\nu-1} < a_i < x_\nu} \frac{1}{a_i} > c_3 \log x_\nu / (\log \log x_\nu)^{\frac{1}{2}}. \tag{5}$$

We shall show that (5) leads to a contradiction and this will prove Theorem 1.

We need the following crucial

LEMMA 1. *Let $u < w \leqq y$, where w is sufficiently large compared to u. Let (the a's are squarefree)*

$$u < a_1 < \cdots < a_k < w, \; a_i \nmid a_j, \qquad 1 \leqq i < j \leqq k \tag{6}$$

and

$$\sum_{i=1}^{k} \frac{1}{a_i} > c_3 \log w / (\log \log w)^{\frac{1}{2}}. \tag{7}$$

Denote by $b_1 < \cdots < b_s \leqq y$ the integers of the form

$$a_i Q_m, \quad Q_m \leqq y / a_i, \qquad 1 \leqq i \leqq k$$

where all the prime factors of Q_m are greater than u. Then

$$\sum_{i=1}^{s} \frac{1}{b_i} > c_4 \log y$$

where c_4 depends only on c_3.

Assume that Lemma 1 has already been proved then we prove Theorem 1 as follows: Let $\lambda c_4 > 2$, $y = x_\lambda$. For each $1 \leqq \nu \leqq \lambda$ we denote by $B^{(\nu)}$ the sequence of integers $b_1^{(\nu)} < \cdots < b_{s_\nu}^{(\nu)}$ of the form

$$a_i Q_m^{(\nu)},\ x_{\nu-1} < a_i < x_\nu, \quad Q_m^{(\nu)} < x_\nu / a_i$$

where all prime factors of $Q_m^{(\nu)}$ are greater than $x_{\nu-1}$. By Lemma 1 we have

$$\sum_{i=1}^{s_\nu} \frac{1}{b_i^{(\nu)}} > c_4 \log y. \tag{8}$$

Now we show that

$$B^{(\nu)} \cap B^{(\nu')} = \emptyset \quad \text{if} \quad \nu \neq \nu'; \tag{9}$$

in other words $b_i^{(\nu)} \neq b_j^{(\nu')}$ if $\nu' > \nu$. If (9) would be false we would have $a_i Q_m^{(\nu)} = a_j Q_n^{(\nu')}$ or $a_i | a_j Q_n^{(\nu')}$. But by our definitions $a_i < x_\nu$, $a_j > x_{\nu'-1} \geqq x_\nu$, thus $a_i < a_j$ and hence $a_i \nmid a_j$. On the other hand all prime factors of $Q_n^{(\nu')}$ are greater than $x_{\nu'-1} \geqq x_\nu$. Thus $(a_i, Q_n^{(\nu')}) = 1$, hence $a_i | a_j Q_n^{(\nu')}$ implies $a_i | a_j$, an evident contradiction. Hence (9) is proved. Clearly $b_i^{(\nu)} < y$ for $\nu \leqq \lambda$. Thus by (8), (9) and $\lambda c_4 > 2$,

$$2 \log y > \sum_{t<y} \frac{1}{t} \geqq \sum_{\nu=1}^{\lambda} \sum_{i=1}^{s_\nu} \frac{1}{b_i^{(\nu)}} \geqq \lambda c_4 \log y > 2 \log y$$

an evident contradiction which proves Theorem 1.

Thus to prove Theorem 1 we only have to prove Lemma 1. We first assume $y = w$ and prove the Lemma in this special case. The general case will follow easily. Denote by $d_1(n)$ the number of divisors of n amongst the a_i, $1 \leqq i \leqq k$, $d_2(n)$ denotes the number of divisors of n amongst the b's. The number of divisors $d(n)$ of the squarefree integer n clearly equals $2^{\nu(n)}$ where $\nu(n)$ is the number of distinct prime factors of n. Clearly

$$\sum_{n=1}^{w} d_2(n) = \sum_{i=1}^{s} \left[\frac{w}{b_i}\right] < w \sum_{i=1}^{s} \frac{1}{b_i}.$$

Thus to prove Lemma 1 in the case $y = w$ it will suffice to show that

$$\sum_{n=1}^{w} d_2(n) > c_5 w \log w. \tag{10}$$

Denote by $n_1 < n_2 < \cdots < n_t < w$ the sequence of integers satisfying

$$\nu(n_i) > \log\log w \quad \text{and} \quad d_1(n_i) > c_6 2^{\nu(n_i)} / \nu(n_i)^{\frac{1}{2}} \tag{11}$$

where c_6 is a sufficiently small constant which will be determined later. Clearly

$$\sum_{i=1}^{t} d_1(n_i) = \sum_{n=1}^{w} d_1(n) - \sum{}' d_1(n) - \sum{}'' d_1(n) \tag{12}$$

where in $\sum'$, $n \leqq w$ and $\nu(n) \leqq \log\log w$, and in $\sum''$

$$n \leqq w,\ \nu(n) > \log\log w,\ d_1(n) \leqq c_6 2^{\nu(n)} / (\nu(n))^{\frac{1}{2}} < c_6 2^{\nu(n)} / (\log\log w)^{\frac{1}{2}}. \tag{13}$$

From (7) we evidently have

$$\sum_{n=1}^{w} d_1(n) \geqq w \sum_{i=1}^{k} \frac{1}{a_i} - w > \frac{c_3}{2} w \log w/(\log\log w)^{\frac{1}{2}}. \tag{14}$$

Clearly

$$\sum{}' d_1(n) < w\, 2^{\log\log w}. \tag{15}$$

From (12) we have

$$\begin{aligned}\sum{}'' d_1(n) &< c_6 \sum_{n=1}^{w} \frac{2^{\nu(n)}}{(\log\log w)^{\frac{1}{2}}} < \frac{c_6}{(\log\log w)^{\frac{1}{2}}} \sum_{n=1}^{w} \frac{w}{n} \\ &< 2c_6 w \log w/(\log\log w)^{\frac{1}{2}}.\end{aligned} \tag{16}$$

From (12), (14), (15) and (16) we have for $c_6 < c_3/10$

$$\sum_{i=1}^{t} d_1(n_i) > \frac{c_3}{4} w \log w/(\log\log w)^{\frac{1}{2}}. \tag{17}$$

Thus to prove (10) we only have to show that for $1 \leqq i \leqq t$

$$d_2(n_i) > c_7 d_1(n_i)\big(\nu(n_i)\big)^{\frac{1}{2}} > c_7 d_1(n_i)(\log\log w)^{\frac{1}{2}}. \tag{18}$$

The last inequality of (18) follows from (11), (17) and (18) clearly imply (10).

To prove (18) let

$$p_1 < \cdots < p_{r_1} \leqq u < q_1 < \cdots < q_{r_2} \leqq w$$

be the prime factors of n_i. Clearly $r_1 < u$, further by (11)

$$r_2 > \log\log w - u > \tfrac{1}{2}\log\log w > r_1$$

if w is sufficiently large (e.g. $w > \exp\exp 2u$). Let $a_1, \cdots, a_l$ be the divisors of n_i amongst the a's. By (11)

$$l > c_6 2^{\nu(n_i)}/\big(\nu(n_i)\big)^{\frac{1}{2}}.$$

To obtain a lower bound for the number of b's dividing n_i, we multiply each a/n_i by all the products of the q's which do not divide a. To show (18) we prove the following combinatorial

Lemma 2. *Let $S = S_1 \cup S_2$, $S_1 \cap S_2 = \emptyset$. The elements of S_1 are $e_1, \cdots, e_k$, the elements of S_2 are $f_1, \cdots, f_l$. Assume $l \geqq k$. Let $A_i \subset S$, $1 \leqq i \leqq r$,*

$$r > c_9\, 2^{k+l}/(k+l)^{\frac{1}{2}} \tag{19}$$

be subsets of S no one of which contains any other. Denote by $B_1, \cdots, B_t$ all the (distinct) subsets of S of the form

$$A_i \cup R, \quad 1 \leqq i \leqq r, \quad R \subset S_2 \tag{20}$$

where in (20) R *runs through all the* 2^l *subsets of* S_2. *Then*

$$t > c_{10} 2^{k+l}.$$

(18) immediately follows from Lemma 2 (to see this it suffices to identify the p's with the e's and the q's with the f's (19) is satisfied because of (11).) Thus to complete the proof of Theorem 1 it suffices to prove Lemma 2.

Before proving Lemma 2 we first need

LEMMA 3. *Let*

$$D_i \subset S_2, \quad 1 \leqq i \leqq j, \quad j > c_1 2^l / l^{\frac{1}{2}}$$

be subsets of the set S_2 *having* l *elements where no* D *contains any other. Let* $E_1, \cdots, E_s$ *be the set of all subsets of* S_2 *which contain at least one* D. *We have*

$$S > c_{14} 2^l.$$

Denote by $\alpha_r \binom{n}{r}$ the number of those D_i for which $|D_i| = r$ and by $\beta_r \binom{n}{r}$ the number of the E_j satisfying $|E| = r$ ($|A|$ is the number of elements of A). We first show

$$\beta_r \geqq \sum_{j \leqq r} \alpha_j. \tag{21}$$

Lemma 3 can be deduced from (21) by a simple computation which we leave to the reader.

To prove (21) it clearly will be sufficient to show that

$$\beta_r \geqq \beta_{r-1} + \alpha_r.$$

Consider all the E's with $|E| = r-1$. Their number is $\beta_{r-1}\binom{n}{r-1}$. Consider now all sets of r elements which contain one of these E's. Their number is $\beta_{r-1}\binom{n}{r-1}(n-r+1)$ and the same set occurs at most r times. Therefore the number of these sets is at least

$$\beta_{r-1}\binom{n}{r-1} \frac{n-r+1}{r} = \beta_{r-1}\binom{n}{r}.$$

These $\beta_{r-1}\binom{n}{r}$ sets are all E's satisfying $|E| = r$ and by assumption none of them are D's having r elements. Hence (21) is evident, and thus Lemma 3 is proved.

We conjectured and Kleitman proved the following stronger result: In a set S of n elements let there be given $\binom{n}{r}$ subsets of S

$$D_1, \cdots, D_{\binom{n}{r}}, \quad D_i \not\subset D_j, \qquad 1 \leqq i < j \leqq \binom{n}{r}.$$

Denote by $E_1, \cdots, E_s$ those subsets of S which contain at least one of the D's. Then

$$s \geqq \sum_{i=0}^{r} \binom{n}{i}.$$

Now we prove Lemma 2. Put

$$A_i = (A_i \cap S_1) \cup (A_i \cap S_2), \qquad 1 \leqq i \leqq r.$$

We split the class of all A's into 2^k classes $C_1, \cdots, C_{2^k}$ where two A's belong to the same class if they have the same intersection with S_1. Let A_{i_1} and A_{i_2} belong to the same class then $A_{i_1} \cap S_2$ clearly does not contain $A_{i_2} \cap S_2$. Hence by the theorem of Sperner [5] each class contains at most

$$\binom{l}{[\frac{1}{2}l]} < c_{11} \frac{2^l}{l^{\frac{1}{2}}} \tag{22}$$

A's. From (22), (19) and $l \geqq k$ we obtain by a simple computation that there are at least $c_{12}\, 2^k$ classes which contain more than $c_{13}\, 2^l/l^{\frac{1}{2}}$ A's. Denote these classes by C_{i_k}, $1 \leqq k \leqq r$, $r > c_{12}\, 2^k$. By Lemma 3 the number of B's for which $B \subset S_1 \cup S_2$ and $B \cap S_2 = A \cap S_2$ where A is in C_{i_k} $(1 \leqq k \leqq r)$ is greater than $c_{14}\, 2^l$. Thus the number of B's is clearly greater than

$$c_{12} c_{14}\, 2^{k+l} > c_{10}\, 2^{k+l}$$

which proves Lemma 2 and therefore Lemma 1 in the case $y = w$.

To prove Lemma 1 in the general case denote by $1 = t_1 < t_2 < \cdots$ the integers all whose prime factors are greater than w. We evidently have

$$\sum_{b_i < y} \frac{1}{b_i} \leqq \sum_{b_i < w} \frac{1}{b_i} \sum_{t_i < y/w} \frac{1}{t_i}. \tag{23}$$

We already proved Lemma 1 if $y = w$, hence

$$\sum_{b_i < w} \frac{1}{b_i} > c_{15} \log w. \tag{24}$$

Further we obtain by a simple computation from a result of de Bruijn [2] that

$$\sum_{t_i < y/w} \frac{1}{t_i} > c_{16} \log y / \log w. \tag{25}$$

Lemma 1 clearly follows from (23), (24) and (25). Thus the proof of Theorem 1 is complete.

It is easy to see that Lemma 2 remains true for $l > c_{16} k$ but fails for $l = o(k)$.

We now state the following sharpening of Theorem 1:

THEOREM 2. *Let A be a primitive sequence, $x_1, x_2, \cdots$ be any sequence satisfying*

$$\log\log x_{\nu+1} > (1+c_{17}) \log\log x_\nu \tag{26}$$

where c_{17} is an arbitrary constant. Put

$$\varepsilon_\nu = \frac{(\log\log x_\nu)^{\frac{1}{2}}}{\log x_\nu} f_A(x_\nu).$$

Then

$$\sum_{\nu=1}^{\infty} \varepsilon_\nu < c_{18}$$

where c_{18} depends only on c_{17}.

We do not give the proof of Theorem 2 since it is very similar to that of Theorem 1 and further (26) can probably be very much improved; perhaps Theorem 2 remains true if (26) is replaced by

$$\log\log x_{\nu+1} > \log\log x_\nu + c_{19} (\log\log x_\nu)^{\frac{1}{2}}.$$

Theorem 1 gives the best upper bound for the growth of $f_A(x)$ for an infinite primitive sequence. Nevertheless further questions can be asked. A well known theorem [4] states that there is an absolute constant c_{20} so that for every primitive sequence,

$$\sum_k \frac{1}{a_k \log a_k} < c_{20}. \tag{27}$$

From (27) we obtain by partial summation

$$\sum_n f_A(2^{2^n})/2^n < c_{21}. \tag{28}$$

Now we prove

THEOREM 3. *Let $g(x)$ be an increasing function for which*

$$\sum_n g(2^{2^n})/2^n = \infty.$$

Then

$$\liminf f_A(x)/g(x) = 0.$$

On the other hand if $g_1(x) = \log x/\log\log x\, h(x)$ where $h(x)$ is increasing and $g_1(x)$ is also increasing and

$$\sum_n g_1(2^{2^n})/2^n \tag{29}$$

converges, then there is a primitive sequence for which

$$\lim f_A(x)/g(x) = \infty. \tag{30}$$

The first part of Theorem 3 follows immediately from (28), hence we only have to prove (29). We will leave some of the details to the reader. Let $p_1 < p_2 < \cdots$ be a sequence of primes for which $\sum 1/p_k < \infty$ and $p_k = (1+o(1))k \log k\, u(k)$ where $u(k) = o(h(k))$. By (29) such a choice is possible. Our primitive sequence consists of the integers of the form

$$p_k t, \quad 1 \leqq k < \infty, \quad \nu(t) = k^2, \quad p_i \nmid t, \quad 1 \leqq i \leqq k.$$

It is not difficult to show by using the methods of [3] that the number of a_i not exceeding x is greater than

$$c_{22} x/u(x) \log \log x.$$

In other words for all sufficiently large n, $a_n < c_{23} n u(n) \log \log n$ or

$$f_A(x) > c_{24} \log x/u(x) \log \log x.$$

In other words (30) holds. The monotonicity conditions on $g(x)$ could no doubt be relaxed, but we do not investigate this question.

References

[1] F. Behrend, 'On sequences of numbers not divisible one by another', *J. London Math. Soc.* 10 (1935), 42—45.

[2] N. G. de Bruijn, 'On the number of uncancelled elements in the sieve of Eratosthenes', *Indagationes Math.* 12 (1950) 247—256.

[3] P. Erdös, 'On the integers having exactly k prime factors', *Annals of Math.* 49 (1948), 53—66.

[4] P. Erdös, 'Note on sequences of integers no one of which is divisible by any other', *J. London Math. Soc.* 10 (1935), 126—128.

[5] E. Sperner, 'Ein Satz über Untermengen einer endlichen Menge', *Math. Zeitschrift* 27 (1928), 544—548.